

(12) **UK Patent Application** (19) **GB** (11) **2 295 011** (13) **A**

(43) Date of A Publication 15.05.1996

(21) Application No **9422687.5**

(22) Date of Filing **10.11.1994**

(71) Applicant(s)
Unilever Plc

(Incorporated in the United Kingdom)

**Unilever House, Blackfriars, LONDON, EC4P 4BQ,
United Kingdom**

(72) Inventor(s)
**Andrew Peter Goldsbrough
Peter Liam Jack**

(74) Agent and/or Address for Service
**B C Roscoe
Unilever Plc, Patent Division, Colworth House,
Sharnbrook, BEDFORD, MK44 1LQ, United Kingdom**

(51) INT CL⁶
C12Q 1/68 , C07H 21/04

(52) UK CL (Edition O)
G1B BAC B104 B108 B121 B200 B203 B224

(56) Documents Cited
EP 0534858 A1 WO 94/01582 A1

(58) Field of Search
UK CL (Edition O) **G1B BAC BAG**
INT CL⁶ **C12Q 1/68**
ONLINE: WPI; BIOTECH; DIALOG

(54) **Selection of DNA markers using adaptor molecules**

(57) A process for the selection of at least one part of a starting DNA which contains a plurality of restriction sites for at least two determined specific restriction endonuclease comprising:

(a) cleaving the starting DNA with a frequent cutting restriction endonuclease and a rare cutting restriction endonuclease with degeneracy associated with the enzyme site to provide a series of restriction endonuclease fragments having a region of overhang;

(b) ligation of the restriction endonuclease fragments to a specific adaptor molecule having a sequence of bases homologous to subsets of the region of overhang to form a tagged restriction endonuclease fragment;

(c) separation of the resultant rare cutting tagged restriction endonuclease fragments.

GB 2 295 011 A

PROCESS FOR GENERATING DNA MARKERSField of Invention

5 This invention relates to a process for generating DNA markers for use in a number of fields including, but not limited to, plant breeding, DNA fingerprinting and most specifically to a method for detecting DNA Markers specific to the genomes of higher plants.

10

Background of the Invention

15 EP 534 858 (Keygene NV) discloses a process for the controlled amplification of at least one part of a starting DNA containing a plurality of restriction sites for a determined specific restriction endonuclease, and of which at least part of its nucleic acid is unknown. The process comprises:

20 (i) digesting the starting DNA with the specific restriction endonuclease or endonucleases, to fragment it into the corresponding series of restriction fragments;

25 (ii) ligating the restriction fragments obtained from the starting DNA with at least one double-stranded synthetic oligonucleotide (adaptor) having one end which is compatible to be ligated to one or both of the ends of the restriction fragments to thereby produce tagged restriction fragments of the starting DNA;

30

(iii) contacting the tagged restriction fragments under hybridizing conditions with at least 1 oligonucleotide primer;

35 (iv) amplifying the tagged restriction fragments hybridized with said primers by PCR or similar techniques; and

(v) identifying or recovering amplified or elongated DNA fragments as produced in step (iv);

5 wherein the primer includes a constant nucleotide sequence which corresponds to the nucleotides involved in the formation of the site for the restriction endonuclease and including at least part of the nucleotides present in the ligated adaptors, and a variable nucleotide sequence, located at the 3' end, which comprises a determined number
10 of nucleotides located immediately adjacent to the last of the nucleotides involved in the restriction site for the endonuclease. Therefore in the process disclosed in EP 534 858 the restriction endonuclease has a constant nucleotide sequence.

15 The selection of tagged restriction fragments is determined by the number of nucleotides residing in the variable sequence part of the primer. The selectivity of the primer increases with the number of nucleotides in the variable
20 (selected) sequence part.

It has been reported that this technology works well with small genome sizes, however problems arise when the technique is used with larger genomes (for example wheat
25 genome).

Selection in EP 534 858 is via the primer, as described above, however such selection may not be 100% precise and polymerisation may still occur from mis-matched sites
30 causing a background amplification. Mis-matching becomes an important problem with large genomes.

We have therefore developed an alternative process for the selection of at least one part of a starting DNA, which
35 process has the advantage that it enables the level of mis-matching to be reduced. Furthermore, in this process the use of PCR amplification techniques are optional and

therefore this new alternative process is greatly simplified.

Disclosure of the Invention

5

Accordingly the invention discloses a process for the selection of at least one part of a starting DNA which contains a plurality of restriction sites for at least two determined specific restriction endonuclease comprising;

10

(a) cleaving the starting DNA with a frequent cutting restriction endonuclease and a rare cutting restriction endonuclease with degeneracy associated with the enzyme site to provide a series of restriction endonuclease fragments having a region of overhang;

15

(b) ligation of the restriction endonuclease fragments to a specific adaptor molecule having a sequence of bases homologous to subsets of the region of overhang to form a tagged restriction endonuclease fragment;

20

(c) separation of the resultant tagged restriction endonuclease fragments;

25

In this specification the following terms are used;

Restriction site - The nucleotide sequence recognised by the restriction endonuclease including the cleavage site. The cleavage site may be within the recognition site or remote from it.

30

Recognition site - The nucleotide sequence recognised by the restriction endonuclease.

35

Degeneracy - The presence of a variable nucleotide sequence located within the restriction site.

Adaptor - short double-stranded DNA molecule with a limited number of base pairs (eg. 10-30 base pairs long) which are designed in such a way that they can be ligated to the sticky end (or overhang region) in the restriction endonuclease fragment.

Rare-cutter - A restriction endonuclease whose specificity is determined by a sequence of >6 bases.

Frequent-cutter restriction endonuclease - a restriction enzyme whose specificity is determined by a sequence of 4 or 5 bases.

The restriction enzymes chosen must provide staggered ends, in which one of the two strands extends beyond the other (commonly known as an overhang or sticky end). Adaptors are used which have single strand extensions which are capable of annealing and ligating to the single strand extensions of the restriction fragments.

Preferably the rare cutter enzyme used is SfiI.

The process of the present invention exploits the use of rare-cutter restriction enzymes with degeneracy associated with the restriction site, to cleave DNA and physically select out a sufficiently small number of genomic fragments to resolve by standard separation techniques. For example SfiI recognises and cleaves the sequence:

GGCCNNNN↓NGGCC
CCGGN↑NNNNCCGG

Assuming a GC content of 50%, then this enzyme will cleave DNA on average about every 64,000 bp. Thus for example, a genome of similar size to the tomato genome (7×10^8 bp) will generate about 2×10^4 SfiI ends. Assuming a random distribution of bases in the degenerate region of the

enzyme site, then any one specific adaptor will anneal to one in 4^3 of these ends i.e. a total of about 300 ends in total. If this specific adaptor is part of an affinity system (e.g. biotinylated) then these 300 fragments can be separated from the rest. In fact this enzyme will cleave much less frequently than this given that the GC content will be less than 50% for most crop species. For a genome of 40% GC the expected number of fragments reduces by a factor of about 6 giving approximately 50 fragments in the above example. Double digestion with a frequent cutting enzyme (e.g. a 4-base recognition site) will result in fragments of length suitable for resolution on standard separation systems (e.g. polyacrylamide gel electrophoresis).

Detection of DNA fragments could involve a number of procedures. For example the biotinylated adaptor could be pre-labelled with radioactive, fluorescent or chromogenic materials or could be so designed to include an internal 3' end for polymerase extension and incorporation of detection substrates. Alternatively, a primer complementary to this adaptor could be annealed to the ligated selected molecules and these subject to linear amplification using DNA polymerase. The adaptor sequence could also be detected by hybridisation with a labelled probe as in conventional RFLP analysis.

Use of different frequent cutting enzymes in the double-digestion with SfiI will generate different fragment profiles for each genome. Comparison of profiles between genotypes (using the same enzyme combinations) will allow restriction fragment polymorphisms to be identified. These can then be used as genetic markers for trait linkage or genetic fingerprinting purposes.

Preferably the tagged restriction endonuclease fragments are separated from the non-tagged restriction endonuclease

fragments via affinity labelling. The specific adaptor is affinity labelled prior to it being ligated to the restriction endonuclease fragment. Suitable affinity systems include biotinylation.

5

Optionally the separated, tagged restriction endonuclease fragment may be contacted with at least one oligonucleotide primer under hybridizing conditions. The fragments hybridised with a primer can then be amplified by polymerase Chain Reaction (PCR). Preferably the primers used are homologous to the adaptor and possibly extend further into a region of degeneracy within the restriction site.

10

Typically, the adaptors used are composed of two synthetic oligonucleotides which are in part complementary to each other, and which are usually approximately 10 to 30 nucleotides long, preferably 12 to 22 nucleotides long and which form double stranded structures when mixed together in solution. Using the enzyme ligase, the adaptors are ligated to the mixture of restriction fragments. When using a large molar excess of adaptors over restriction fragments one ensures that all restriction fragments will end up carrying adaptors at both ends.

20

The process of the invention provides a means to permit the detection between source samples of polymorphisms caused both by length differences (resulting from deletions, additions, inversions) which have resulted in either loss or gain of a restriction site, or changes of the nucleotide sequence in either the recognition or cleavage site of an enzyme which recognises and cleaves such sites. The invention includes methods for detecting these polymorphisms, synthetic oligonucleotides for use in the methods of the invention, applications of the methods and procedures of the invention in a number of fields including plant breeding, and DNA fingerprinting.

30

35

Specifically, the methods described here provide an alternative means of identifying genomic restriction fragments which are either genetically linked to one or more particular traits, or which can provide a fingerprint of the genome under examination.

The present invention is based on the definition of novel specific methods to achieve selectivity of restriction fragments. These selected restriction fragments may optionally be subsequently amplified.

In general, restriction endonuclease digests of genomic DNA, and in particular of plant genomic DNA, yields very large numbers of restriction fragments, the exact number depending upon the size of the genome and the frequency of occurrence of the recognition site of the restriction endonuclease in the genome, which in turn is primarily determined by the number of nucleotides in the recognition sequence - a number typically ranging between 4 and 8. Generally the number of restriction fragments produced is too large to enable identification of individual fragments fractionated by gel electrophoresis or other fractionation methods.

We have used a novel method to limit the number of restriction fragments which may optionally be subsequently amplified. The basis for selection resides in the choice of restriction endonuclease used to digest the genomic DNA, and in the design of adaptor oligonucleotide. The selective principle resides in the use of restriction endonucleases which have within their cleavage and recognition sites a number of completely degenerate nucleotides. Selection is determined by the number of specific nucleotides in the terminal extensions of the adaptor.

It is possible to estimate the degree of selectivity

obtainable by the selective adaptors using the general formula 4^n , where n equals the number selective bases present in the adaptor (assuming all nucleotides are represented equally within the degenerate region). Thus, when 1 selective base is used (the terminal base of the extension), 1 out of every 16 fragments will be capable of annealing to and ligating with the (partially) specific adaptor. Using 2 selective bases (the terminal 2 bases of the adaptor extension) 1 out of every 256 fragments will be capable of annealing to and ligating with the adaptor; using 3 selective bases (the terminal 3 bases of the adaptor extension) 1 out of every 4096 fragments will be capable of annealing to and ligating with the adaptor; using 4 selective bases (the terminal 4 bases of the adaptor extension) 1 out of every 65,536 fragments will be capable of annealing to and ligating with the adaptor; and so on. Other combinations with some completely selective and some partially selective (i.e. partially degenerate) bases give rise to different numbers of compatible ligation reactions.

The products obtained in accordance with the invention can be identified using standard fractionation techniques known to those skilled in the art using, for example but not limited to, agarose or acrylamide gel electrophoresis. The invention permits the number of products obtained to be tuned in accordance with the resolution of the fractionation system being employed. Products may be visualised directly following staining of the molecules with appropriate agents. Alternatively, the primers or nucleotides present for the PCR amplification (if used) may be labelled with radioactivity or a fluorescent chromophore, thus allowing identification of reaction products after size fractionation.

In accordance with the invention, different sets of amplified products are obtained with the different sets of selective adaptors. The banding patterns identified

following fractionation constitute unique and reproducible fingerprints of the genomic DNA. Such fingerprints can have several uses such as, but not limited to, forensic typing, diagnostic identification of organisms, and the
5 identification of species, varieties or individuals. The level of identification will be determined by the degrees of similarity and differences between the members of the group being studied. The underlying principle of the invention is that in each product two nucleotide sequences
10 are detected (the target restriction site) which are separated from each other by a given distance. In related organisms, species, varieties, races or individuals these two sequences and the relative distances separating them will be conserved to a greater or lesser degree. Hence the
15 fingerprints obtained constitute a basis for determining the degree of sequence relationships between genomes. The fingerprints can also be used to distinguish genomes from each other.

20 Another particular application of the invention involves the screening and identification of restriction fragment length polymorphisms (RFLPs). Changes in the nucleotide composition of genomic DNA can often result in polymorphisms of restriction fragments: insertions or
25 deletions affect size of the fragments containing them; nucleotide changes can result in the elimination or creation of new endonuclease target recognition sites. Restriction fragment polymorphisms of this nature can be identified directly by comparing products from different
30 genomes.

RFLPs are particularly useful for monitoring the inheritance of agronomic traits in plant breeding, in that certain DNA polymorphisms which are closely linked with
35 specific genetic traits can be used to monitor for the presence or absence of the said trait.

The present invention provides a general method for isolating DNA markers from any genome and for using such DNA markers in all possible applications of DNA fingerprinting.

Claims

1. A process for the selection of at least one part of a starting DNA which contains a plurality of restriction sites for at least two determined specific restriction endonuclease comprising:

(a) cleaving the starting DNA with a frequent cutting restriction endonuclease and a rare cutting restriction endonuclease with degeneracy associated with the enzyme site to provide a series of restriction endonuclease fragments having a region of overhang;

(b) ligation of the restriction endonuclease fragments to a specific adaptor molecule having a sequence of bases homologous to subsets of the region of overhang to form a tagged restriction endonuclease fragment;

(c) separation of the resultant rare cutting tagged restriction endonuclease fragments;

2. A process according to claim 1 wherein the adaptor is affinity labelled and the subset of restriction endonuclease fragments which have the adaptor ligated to them are thus separated.

3. A process according to claim 1 or 2 wherein the rare cutting restriction endonuclease is SfiI.

4. A process according to any preceding claim wherein the tagged restriction endonuclease is contacted with at least one oligonucleotide primer under hybridizing conditions; the fragments hybridised with a primer are then amplified by Polymerase Chain Reaction (PCR).



The Patent Office

-12-

Application No: GB 9422687.5
Claims searched: 1-4

Examiner: Dr. N.R. Curtis
Date of search: 30 January 1996

Patents Act 1977 Search Report under Section 17

Databases searched:

UK Patent Office collections, including GB, EP, WO & US patent specifications, in:

UK CI (Ed.O): G1B (BAC, BAG)

Int CI (Ed.6): C12Q 1/68

Other: ONLINE: WPI; BIOTECH/DIALOG

Documents considered to be relevant:

Category	Identity of document and relevant passage	Relevant to claims
A	EP 0,534,858 A1 (Keygene N.V.) (See Example 1)	1,2,4
A	WO 94/01582 A1 (Medical Research Council) (See page 11, lines 23-26; page 13, line 8 to page 14, line 1).	1,2,4

X Document indicating lack of novelty or inventive step
Y Document indicating lack of inventive step if combined with one or more other documents of same category.
& Member of the same patent family

A Document indicating technological background and/or state of the art.
P Document published on or after the declared priority date but before the filing date of this invention.
E Patent document published on or after, but with priority date earlier than, the filing date of this application.